

УДК 519.237.5

## ДИНАМИЧЕСКОЕ РЕГРЕССИОННОЕ МОДЕЛИРОВАНИЕ НА ОСНОВЕ ГРАДИЕНТНОГО СПУСКА ПО УЗЛОВЫМ ПРЯМЫМ

<sup>1,2</sup>Тырсин А.Н., <sup>1</sup>Голованов О.А.

<sup>1</sup>ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Екатеринбург, e-mail: at2001@yandex.ru;

<sup>2</sup>ФГБУН Научно-инженерный центр «Надежность и ресурс больших систем и машин» УрО РАН, Екатеринбург, e-mail: at2001@yandex.ru

Использование регрессионного анализа в задачах непрерывного мониторинга состояния систем требует высокого быстродействия алгоритма оценивания параметров модели. При этом метод оценивания должен обладать устойчивостью к стохастической неоднородности данных. Такому условию соответствует метод наименьших модулей. В работе рассмотрен один из наиболее быстрых вариантов его реализации, основанный на спуске по узловым прямым. Целью статьи является повышение вычислительной эффективности известного метода на основе спуска по узловым прямым. В статье описан алгоритм линейного регрессионного моделирования многомерных динамических процессов на основе метода наименьших модулей. Алгоритм реализует градиентный спуск по узловым прямым. Вычислительная сложность оценивания коэффициентов регрессии в динамике снижена за счет использования в качестве начальной точки решения задачи на предыдущем шаге. Выигрыш достигается за счет снижения количества переходов с одной узловой прямой на другую. Это количество не зависит от объема выборки. Выполнена оценка вычислительной сложности разработанного алгоритма. Проведен сравнительный анализ вычислительной эффективности предложенного и известного алгоритмов. Также показано, что наилучшей стратегией спуска из узловой точки является рассмотрение всех узловых прямых, проходящих через эту точку. Разработанный алгоритм также позволит повысить быстродействие реализации обобщенного метода наименьших модулей.

**Ключевые слова:** метод наименьших модулей, линейная регрессия, динамика, алгоритм, вычислительная эффективность, узловая прямая

## DYNAMIC REGRESSION MODELING BASED ON GRADIENT DESCENT ALONG NODAL STRAIGHT LINES

<sup>1,2</sup>Tyrsin A.N., <sup>1</sup>Golovanov O.A.

<sup>1</sup>Ural Federal University named after the first President of Russia B.N. Yeltsin, Yekaterinburg, e-mail: at2001@yandex.ru;

<sup>2</sup>Federal State Budgetary Institution of Science Scientific-Engineering Center Reliability and Life of Large Systems and Machines, Ural Branch, Russian Academy of Science, Yekaterinburg, e-mail: at2001@yandex.ru

The use of regression analysis in the tasks of continuous monitoring of the state of systems requires a high speed of the algorithm for estimating the parameters of the model. At the same time, the estimation method must be resistant to stochastic heterogeneity of the data. This condition corresponds to the method of least modules. The paper considers one of the fastest variants of its implementation, based on the descent along the nodal straight lines. The purpose of the article is to increase the computational efficiency of the known method based on descent along nodal straight lines. The article describes an algorithm for linear regression modeling of multidimensional dynamic processes based on the least absolute deviations method. The algorithm implements gradient descent along nodal straight lines. The computational complexity of estimating regression coefficients in dynamics is reduced by using the problem solution as the starting point at the previous step. The gain is achieved by reducing the number of transitions from one node line to another. This number does not depend on the sample size. The computational complexity of the developed algorithm is estimated. A comparative analysis of the computational efficiency of the proposed and known algorithms is carried out. It is also shown that the best strategy for descending from a nodal point is to consider all nodal lines passing through this point. The developed algorithm will also improve the performance of the implementation of the generalized least absolute deviations method.

**Keywords:** least absolute deviations method, linear regression, dynamics, algorithm, computational efficiency, nodal straight line

Линейная многомерная регрессия – систем [1–3]. Для ее использования по наблюдаемым данным ( $\mathbf{X}$ ,  $\mathbf{y}$ ) оценивают вектор  $\mathbf{a}$  коэффициентов в модели

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \boldsymbol{\varepsilon}, \quad (1)$$

где  $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$ ,  $\mathbf{X} = \{x_{ij}\}_{n \times m} = \begin{pmatrix} 1 & x_{12} & \dots & x_{1m} \\ 1 & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n2} & \dots & x_{nm} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \dots \\ \mathbf{x}_n \end{pmatrix}$ ,  $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_m \end{pmatrix}$ ,  $\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$  – вектор нена-

блюдаемых случайных погрешностей (ошибок, невязок).

Наиболее распространенным методом оценивания параметров регрессионных моделей по экспериментальным данным является метод наименьших квадратов (МНК). Однако данный метод требует выполнения ряда предпосылок, называемых условиями Гаусса – Маркова [4, 5], и при их нарушении может приводить к значительному снижению достоверности моделирования. В этой ситуации применяют устойчивые методы, в первую очередь основанные на минимизации суммы модулей (или функций от модулей) невязок. В первом случае используется метод наименьших модулей (МНМ) [6, 7]. МНМ для нахождения коэффициентов регрессии в (1) представляет собой задачу минимизации

$$Q(\mathbf{a}) = \sum_{i=1}^n |y_i - \langle \mathbf{a}, \mathbf{x}_i \rangle| \rightarrow \min_{\mathbf{a} \in \mathbf{R}^m}. \quad (2)$$

Для обеспечения большей устойчивости можно применять обобщенный метод наименьших модулей (ОМНМ) [8]. ОМНМ-оценки для (1) находят как решение задачи

$$W(\mathbf{a}) = \sum_{i=1}^n \rho_i(|y_i - \langle \mathbf{a}, \mathbf{x}_i \rangle|) \rightarrow \min_{\mathbf{a} \in \mathbf{R}^m}, \quad (3)$$

где  $\rho(\cdot)$  – некоторая монотонно возрастающая, дважды непрерывно-дифференцируемая на положительной полуоси функция, причем  $\rho(0) = 0$ ,  $\forall t > 0 \quad 0 < \rho'(t) < \infty$ ,  $-\infty < \rho''(t) < 0$ .

МНМ и ОМНМ объединяет общее свойство – в обоих случаях решение находится в узловых точках. В [9, 10] был предложен новый подход к решению задач (2) и (3) на основе спуска по узловым прямым. Так средняя вычислительная сложность алгоритма градиентного спуска по узловым прямым для реализации МНМ составила  $O(m^2 n^{1.4})$  [10], а вычислительная сложность нахождения ОМНМ-оценки –  $O(m^3 n^2 \ln n + C_{\alpha, \beta}^m \cdot (mn + m^3))$ , где  $\alpha$  – доля гиперплоскостей, ближайших к текущей узловой точке, рассматриваемых на каждой итерации [9]. Несмотря на достаточно высокое быстродействие, вычислительная эффективность данных алгоритмов может оказаться недостаточной для решения задач мониторинга состояния систем, в которых

оценивание параметров модели происходит в режиме реального времени в темпе поступления новых входных данных.

Цель исследования – повышение вычислительной эффективности алгоритмов спуска по узловым прямым для решения динамических задач мониторинга состояния систем. Рассмотрим обе задачи – (2) и (3).

### Материалы и методы исследования

Введем гиперплоскости  $\Omega_i = \Omega(\mathbf{a}, \mathbf{x}_i, y_i)$  в виде уравнений

$$y_i - \langle \mathbf{a}, \mathbf{x}_i \rangle = 0, \quad (i = 1, 2, \dots, n). \quad (4)$$

Зададим также узловые точки пересечения гиперплоскостей (4)

$$\mathbf{u}_{(k_1, \dots, k_m)} = \bigcap_{i=k_1}^{k_m} \Omega_i,$$

$$1 \leq k_1 < k_2 < \dots < k_m \leq n, \quad k_s \in N. \quad (5)$$

Будем называть узловой прямой, являющуюся пересечением  $(m - 1)$  непараллельных гиперплоскостей  $\Omega_i$ :

$$l_{(k_1, \dots, k_{m-1})} = \bigcap \Omega_i, \quad i \in \{k_1, k_2, \dots, k_{m-1}\}. \quad (6)$$

Обозначим:  $\Omega = \{\Omega_1, \Omega_2, \dots, \Omega_n\}$  – множество всех гиперплоскостей,  $U$  – множество всех узловых точек (5). Минимум целевых функций  $Q(\mathbf{a})$  и  $W(\mathbf{a})$  всегда принадлежит множеству  $U$ .

Алгоритм решения задачи (2) основан на спуске к точному решению, двигаясь вдоль узловых прямых (6). В [9] показано, что минимум функции  $Q(\mathbf{a})$  принадлежит множеству  $U$  и достигается за конечное число шагов.

За начальное приближение берем узловую точку  $\mathbf{u}^{(0)} = \mathbf{u}_{(k_1, \dots, k_m)}$ , являющуюся пересечением  $m$  произвольных гиперплоскостей  $\Omega_{k_1}, \Omega_{k_2}, \dots, \Omega_{k_m}$ , путем решения системы линейных алгебраических уравнений (СЛАУ):  $y_i - \langle \mathbf{u}^{(0)}, \mathbf{x}_i \rangle = 0$ ,  $i = \{k_1, k_2, \dots, k_m\}$ . Например, можно взять  $i = 1, 2, \dots, m$ . Исключив одну из гиперплоскостей, например  $\Omega_{k_m}$ , получим узловую прямую  $l_{(k_1, \dots, k_{m-1})}$ . Расширенная матрица СЛАУ прямой  $l_{(k_1, \dots, k_{m-1})}$  имеет вид

$$\mathbf{A}_{(k_1, \dots, k_{m-1})} = \begin{pmatrix} 1 & x_{k_1,2} & x_{k_1,3} & \dots & x_{k_1,m-1} & x_{k_1,m} & y_{k_1} \\ 1 & x_{k_2,2} & x_{k_2,3} & \dots & x_{k_2,m-1} & x_{k_2,m} & y_{k_2} \\ 1 & x_{k_3,2} & x_{k_3,3} & \dots & x_{k_3,m-1} & x_{k_3,m} & y_{k_3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{k_{m-1},2} & x_{k_{m-1},3} & \dots & x_{k_{m-1},m-1} & x_{k_{m-1},m} & y_{k_{m-1}} \end{pmatrix}.$$

Применив алгоритм прямого хода метода Гаусса, преобразуем матрицу  $\mathbf{A}_{(k_1, \dots, k_{m-1})}$  к ступенчатому виду

$$\mathbf{A}'_{(k_1, \dots, k_{m-1})} = \begin{pmatrix} 1 & x_{k_1,2} & x_{k_1,3} & \dots & x_{k_1,m-1} & x_{k_1,m} & y_{k_1} \\ 0 & 1 & x'_{k_2,3} & \dots & x'_{k_2,m-1} & x'_{k_2,m} & y'_{k_2} \\ 0 & 0 & 1 & \dots & x'_{k_3,m-1} & x'_{k_3,m} & y'_{k_3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & x'_{k_{m-1},m} & y'_{k_{m-1}} \end{pmatrix}.$$

Используя ступенчатую матрицу  $\mathbf{A}'_{(k_1, \dots, k_{m-1})}$ , можно значительно сократить вычислительные затраты на нахождение всех узловых точек, лежащих на прямой  $l_{(k_1, \dots, k_{m-1})}$ . Действительно, для любой  $i$ -й узловой точки имеем расширенную матрицу

$$\mathbf{A}_{(k_1, \dots, k_{m-1}, i)} = \begin{pmatrix} 1 & x_{k_1,2} & x_{k_1,3} & \dots & x_{k_1,m-1} & x_{k_1,m} & y_{k_1} \\ 0 & 1 & x'_{k_2,3} & \dots & x'_{k_2,m-1} & x'_{k_2,m} & y'_{k_2} \\ 0 & 0 & 1 & \dots & x'_{k_3,m-1} & x'_{k_3,m} & y'_{k_3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & x'_{k_{m-1},m} & y'_{k_{m-1}} \\ 1 & x_{i,2} & x_{i,3} & \dots & x_{i,m-1} & x_{i,m} & y_i \end{pmatrix}, \quad (7)$$

где  $1 \leq k_1 < k_2 < \dots < k_{m-1} \leq n$ ,  $i \in \{1, 2, \dots, n\}$ ,  $i \notin \{k_1, k_2, \dots, k_{m-1}\}$ .

Находим все узловые точки, лежащие на  $l_{(k_1, \dots, k_{m-1})}$ . Варьируя номер  $i$  в (7), найдем все  $(n - 1)$  узловых точек на прямой и упорядочиваем их по направлению этой прямой. Далее в алгоритмах спуска по узловым прямым в одной из узловых точек  $\mathbf{u}_{(k_1, \dots, k_{m-1}, j)}$  с помощью направленного (градиентного) спуска минимум функции  $Q(\mathbf{a})$  на прямой  $l_{(k_1, \dots, k_{m-1})}$ .

Линейный вид выражений под знаком модуля в (2) приводит к простому виду производной функции  $Q(\mathbf{a})$  по направлению узловой прямой. Покажем это.

В узловой точке  $\mathbf{u}^{(0)} = \mathbf{u}_{(k_1, \dots, k_m)}$  вычисляем значение производной функции  $Q(\mathbf{a})$  по направлению узловой прямой  $l_{(k_1, \dots, k_{m-1})}$ . Двигаемся из узловой точки  $\mathbf{u}^{(0)}$  по направлению убывания производной по направлению. Эта производная не существует только в узловых точках. Поэтому в очередной узловой точке  $\mathbf{u}^{(1)} = (u_1^{(1)}, u_2^{(1)}, \dots, u_m^{(1)})^T$  находим производную с двух сторон. Каждая из производных равна сумме  $n$  слагаемых по числу наблюдений [10]

$$\frac{\partial Q(\mathbf{u}^{(1)})}{\partial \mathbf{l}_{(k_1, \dots, k_{m-1})}} = \sum_{i=1}^n (c_1 + x_{i2}c_2 + \dots + x_{im}c_m) \cdot \text{sign}\left(\sum_{j=1}^m u_j^{(1)} x_{ij} - y_i\right). \quad (8)$$

где  $\mathbf{l}_{(k_1, \dots, k_{m-1})} = (c_1, c_2, \dots, c_m)^T$  – направляющий вектор узловой прямой  $l_{(k_1, \dots, k_{m-1})}$ . При переходе через узловую точку изменяется только одно из слагаемых в (8). Его номер соответствует номеру гиперплоскости, пересекающей в этой узловой точке  $\mathbf{u}^{(1)}$  прямую  $l_{(k_1, \dots, k_{m-1})}$ . Если

знак производной по направлению перед  $\mathbf{u}^{(1)} \left. \frac{\partial Q}{\partial \mathbf{l}_{(k_1, \dots, k_{m-1})}} \right|_{\mathbf{u}^{(1)}}$  и после  $\mathbf{u}^{(1)} \left. \frac{\partial Q}{\partial \mathbf{l}_{(k_1, \dots, k_{m-1})}} \right|_{\mathbf{u}^{(1)}}$  меня-

ется с отрицательного на положительной (или производная станет равной нулю), то в этой точке целевая функция принимает наименьшее значение по узловой прямой. Тогда вместо  $\mathbf{u}^{(0)}$  фиксируем  $\mathbf{u}^{(1)}$ . Если на данной прямо целевая функция в исходной узловой точке  $\mathbf{u}^{(0)}$  принимает наименьшее значение (производная по направлению при переходе через  $\mathbf{u}^{(0)}$  в ту и другую сторону становится больше или равной нулю), то переходим к следующей узловой прямой, проходящей через эту узловую точку.

В случае, когда вместо  $\mathbf{u}^{(0)}$  фиксируем  $\mathbf{u}^{(1)}$ , тогда из точки  $\mathbf{u}^{(1)}$  переходим на другую узловую прямую и продолжаем спуск по ней по тому же принципу. В результате будет найдена узловая точка  $\mathbf{u}^{(*)}$ , спуск из которой невозможен. Эта узловая точка будет являться точным решением задачи (2). В [10] рассмотрены два варианта продолжения спуска после нахождения очередной узловой точки:

1) продолжить спуск из точки  $\mathbf{u}_{(k_1, \dots, k_{m-1}, j)}$  по другой узловой прямой и т.д. до тех пор, пока спуск будет невозможен;

2) выполнить спуск из точки  $\mathbf{u}_{(k_1, \dots, k_{m-1}, j)}$  по всем остальным  $(m - 1)$  узловым прямым, найти в каждой из них минимум  $Q(\mathbf{a})$  и перейти в узловую точку с наименьшим значением целевой функции и снова продолжить спуск из этой точки.

Однако остался открытым вопрос совмещения этих вариантов проверки узловых прямых: проведение какой-то «части» итераций с использованием одного из способов проверки узловых прямых, остальные же итерации будут производиться другим способом. Итерацией же, относительно поставленной задачи, будет считаться переход от одной узловой точки к другой.

### Результаты исследования и их обсуждение

Проверим гипотезу о совмещении двух описанных ранее способов проверки узловых прямых. Данная гипотеза заключается в проведении какой-то «части» итераций с использованием одного из способов проверки узловых прямых, остальные же итерации будут производиться другим способом. Итерацией же, относительно поставленной задачи, будет считаться переход от одной узловой точки к другой.

Количество итераций, для последующего деления на «части», будет вычисляться алгоритмом при анализе выборки с использованием способа проверки узловых прямых без совмещения. К примеру, в первую очередь выборка будет проанализирована способом с полной проверкой всех прилегающих узловых прямых. Тогда в последующем, взяв количество итераций с предыдущего шага, первая «часть» итераций (в нашем случае 10%, ..., 90%) будет осу-

ществлена способом с полной проверкой прямых, а все остальные – способом с переходом на первую узловую прямую с меньшим значением целевой функции.

С помощью метода статистических испытаний Монте-Карло [11, 12] было проведено исследование, в котором для  $m = 2, 3, \dots, 10$  оценивалось время решения задачи (2). Объемы выборок  $n = 50, 100, \dots, 500$ . Результаты приведены в табл. 1.

В итоге наилучшие результаты показал способ *Full*. Отметим, что подобная тенденция в скорости вычисления сохраняется как при нормальном распределении выборки, так и при наличии выбросов. Поэтому наибольшим быстродействием обладает спуск из узловой точки  $\mathbf{u}_{(k_1, \dots, k_{m-1}, j)}$  по всем  $m$  узловым прямым.

Анализ данных показал, что в этом случае среднее число рассмотренных узловых точек имеет линейную зависимость с объемом выборки  $n$  и степенную с количеством коэффициентов в наблюдении  $m$ . Среднее число рассмотренных узловых точек достаточно точно аппроксимируется зависимостью  $L(m, n) = O(m^2n)$ . Вычислительная сложность решения задачи (3) может быть снижена за счет более быстрого нахождения начальной точки, являющейся решением задачи (2).

Рассмотрим теперь динамическую задачу. Пусть мы нашли решение  $\mathbf{a}^{(*)}$  задачи (2) по начальной выборке данных  $(\mathbf{X}_i, y_i) = (x_{i1}, \dots, x_{im}, y_i)$ ,  $i = 1, \dots, n$ . Теперь в динамике добавляем новое наблюдение  $(\mathbf{X}_{n+1}, y_{n+1})$ , удаляем самое «старое» первое наблюдение  $(\mathbf{X}_1, y_1)$  и снова оцениваем МНМ-параметры уравнения регрессии.

Таблица 1

Доля наиболее быстрых решений задачи (2) различными способами проверки узловых прямых

Способ проверки прямой	% от общего числа опытов	Способ проверки прямой	% от общего числа опытов
First	15,00	Full	39,72
10%	4,44	10%	1,94
20%	3,89	20%	1,39
30%	4,72	30%	1,67
40%	3,33	40%	1,94
50%	3,06	50%	1,11
60%	3,61	60%	1,39
70%	4,17	70%	0,83
80%	2,78	80%	0,56
90%	3,61	90%	0,83

Примечание. *Full* – способ с проверкой всех прилегающих узловых прямых; *First* – способ с переходом на первую узловую прямую с меньшим значением целевой функции относительно исходной точки.

Решение  $\mathbf{a}^{(*)}$  представляет собой пересечение  $m$  гиперплоскостей (обозначим их как множество  $G_1$ ), образованных некоторыми  $m$  наблюдениями (обозначим их как множество  $Z_1$ ). Возможны два случая. Во-первых, наблюдение  $(\mathbf{X}_1, y_1)$  не принадлежит к множеству  $Z_1$ . Тогда считаем точку  $\mathbf{a}^{(*)}$  начальной точкой (присваиваем  $\mathbf{a}^{(0)} = \mathbf{a}^{(*)}$ ) и начинаем искать решение по алгоритму градиентного спуска по узловым прямым.

Во-вторых, наблюдение  $(\mathbf{X}_1, y_1)$  принадлежит множеству  $Z_1$ . Добавляем к множеству  $Z_1 \setminus (\mathbf{X}_1, y_1)$  любое из не принадлежащих ему наблюдений, например  $(\mathbf{X}_{n+1}, y_{n+1})$ , находим пересечение оставшихся  $(m-1)$  гиперплоскостей из множества  $G_1$  (их пересечение образует узловую прямую  $l_{(k_1^0, \dots, k_{m-1}^0)}$ ) с гиперплоскостью  $\Omega_{n+1}$ , образованной наблюдением  $(\mathbf{X}_{n+1}, y_{n+1})$ . Отметим, что гиперплоскость  $\Omega_{n+1}$  и узловая прямая  $l_{(k_1^0, \dots, k_{m-1}^0)}$  должны пересекаться (не должны быть параллельными). Тогда точку  $\mathbf{a}^{(0)} = l_{(k_1^0, \dots, k_{m-1}^0)} \cap \Omega_{n+1}$  считаем начальной точкой и начинаем искать решение по алгоритму градиентного спуска по узловым прямым.

Выигрыш в быстродействии в динамике должен достигаться за счет более близкого расположения узловой точки  $\mathbf{a}^{(0)}$  к решению задачи (2). В [13] получено, что среднее число переходов от одной узловой прямой на другую составляет  $O(m \ln n)$ . Учет полученного на предыдущем шаге решения в динамической задаче должен уменьшить

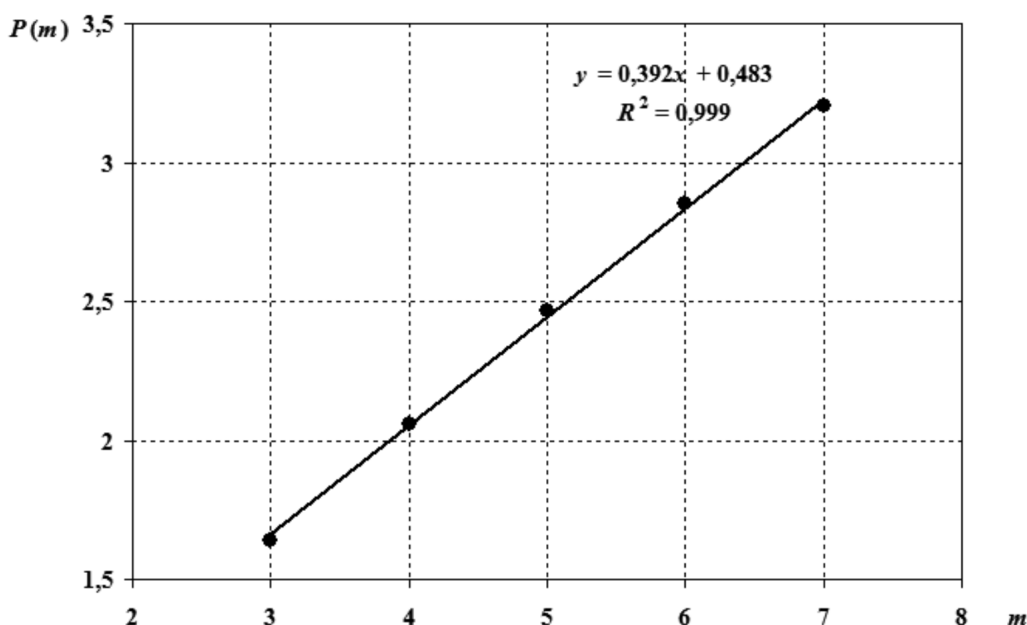
это число. Снова воспользуемся методом Монте-Карло. В табл. 2 для 2000 испытаний приведены средние значения  $p(m, n)$  общего количества переходов с одной узловой прямой на другую, для тех же значений  $m$  и  $n$ , что и в [13].

Таблица 2

Средние значения количества переходов с одной узловой прямой на другую

$n$	$p(m, n)$				
	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$
30	1,64	2,09	2,49	2,81	3,02
50	1,66	2,01	2,44	2,79	3,11
100	1,70	2,05	2,44	2,82	3,18
150	1,63	2,06	2,44	2,82	3,26
300	1,65	2,07	2,40	2,89	3,19
500	1,62	2,07	2,53	2,94	3,28
700	1,65	2,10	2,51	2,86	3,33
900	1,58	2,05	2,46	2,84	3,23
1000	1,66	2,05	2,50	2,87	3,18
1200	1,65	2,09	2,50	2,82	3,24
1500	1,63	1,97	2,43	2,89	3,23
1700	1,67	2,07	2,46	2,84	3,25
1850	1,61	2,03	2,43	2,92	3,20
2000	1,63	2,08	2,49	2,81	3,18
Среднее $P(m)$	1,64	2,06	2,47	2,85	3,21

Видим, что  $p(m, n)$  практически не зависят от объема выборки  $n$ , а средние значения  $P(m)$  линейно зависят только от числа параметров  $m$  регрессионной зависимости (рисунок). На рисунке показана линейная регрессионная зависимость  $P(m)$  от  $m$ .



Линейная регрессионная зависимость  $P(m)$  от числа параметров  $m$



Для сравнения в табл. 3 приведем отношения средних значений общего количества переходов с одной узловой прямой на другую в статическом случае  $p_0(m, n)$  (значения взяты из [13]) и в динамическом  $p(m, n)$  (табл. 2).

**Таблица 3**

Отношения  $p_0(m, n)/p(m, n)$  средних значений общего количества переходов с одной узловой прямой на другую в статическом и в динамическом случаях

n	$p_0(m, n)/p(m, n)$				
	m = 3	m = 4	m = 5	m = 6	m = 7
30	3,20	3,18	3,11	3,14	3,19
50	3,69	3,97	4,01	4,03	4,07
100	4,33	4,82	5,05	5,05	5,29
150	4,83	5,22	5,62	5,79	5,85
300	5,55	6,11	6,71	6,80	7,16
500	6,06	6,71	7,11	7,40	7,93
700	6,34	7,15	7,61	8,21	8,45
900	6,88	7,55	8,10	8,59	9,07
1000	6,63	7,68	8,14	8,70	9,46
1200	6,84	7,76	8,41	9,23	9,63
1500	7,23	8,51	8,93	9,31	10,04
1700	7,15	8,28	8,97	9,74	10,20
1850	7,48	8,54	9,26	9,59	10,47
2000	7,47	8,48	9,17	10,05	10,62

Таким образом, получили, что среднее число переходов от одной узловой прямой на другую в предложенном динамическом варианте спуска по узловым прямым составляет  $O(m)$ . С учетом [10] средняя вычислительная сложность динамического варианта градиентного спуска по узловым прямым составит  $L = O(m^2 n \cdot \max(m^2, \ln n))$ . Если  $n \gg m$ , то  $L = O(m^2 n \cdot \max(m^2, \ln n)) \leq O(m^2 n \ln n)$ , т.е. выигрыш составил примерно в  $O(\ln n)$  раз.

### Заключение

1. Предложенный алгоритм реализации спуска по узловым прямым позволяет повысить быстродействие при решении задач регрессионного моделирования в динамике примерно в  $O(\ln n)$  раз.

2. Поскольку после удаления на каждом шаге одного наблюдения узловые прямые почти не меняются, следует в дальнейшем также учесть эту особенность рассматриваемой задачи для повышения быстродействия алгоритма.

3. Разработанный алгоритм также позволит повысить быстродействие реализации обобщенного метода наименьших модулей.

*Работа выполнена при финансовой поддержке гранта РФФИ, проект № 20-41-660008 p\_a.*

### Список литературы

1. Darlington R.B., Hayes A.F. Regression Analysis and Linear Models. Concepts, Applications, and Implementation. The Guilford Press, 2017. 688 p.
2. Gunst R.F., Mason R.L. Regression Analysis and Its Application. A Data-Oriented Approach. CRC Press, 2017. 422 p.
3. Тюрин Ю.Н. Многомерная статистика: гауссовские линейные модели. М.: Издательство Московского университета, 2011. 136 с.
4. Демиденко Е.З. Линейная и нелинейная регрессия. М.: Финансы и статистика, 1981. 302 с.
5. Дронов С.В. Методы и задачи многомерной статистики. Барнаул: Изд-во Алтайского государственного университета, 2015. 276 с.
6. Мудров В.И., Кушко В.Л. Методы обработки измерений. Квазиправдоподобные оценки. М.: Радио и связь, 1983. 304 с.
7. Тырсин А.Н., Максимов К.Е. Оценивание линейных регрессионных уравнений с помощью метода наименьших модулей // Заводская лаборатория. Диагностика материалов. 2012. Т. 78. № 7. С. 65–71.
8. Тырсин А.Н., Соколов Л.А. Оценивание линейной регрессии на основе обобщенного метода наименьших модулей // Вестник Самарского государственного технического университета. Сер. Физ.-мат. науки. 2010. № 5 (21). С. 134–142.
9. Тырсин А.Н., Азарян А.А. Методы устойчивого построения линейных моделей на основе спуска по узловым прямым // Модели, системы, сети в экономике, технике, природе и обществе. 2018. № 1 (25). С. 188–202.
10. Тырсин А.Н. Алгоритмы спуска по узловым прямым в задаче оценивания регрессионных уравнений методом наименьших модулей // Заводская лаборатория. Диагностика материалов. 2021. Т. 87. № 5. С. 68–75.
11. Михайлов Г.А., Войтишек А.В. Численное статистическое моделирование. Методы Монте-Карло. Издательский центр «Академия», 2006. 368 с.
12. Kroese D.R., Taimre Th., Botev Z.I. Handbook of Monte Carlo Methods. Wiley, 2011. 774 p.
13. Азарян А.А. Быстрые алгоритмы моделирования многомерных регрессионных зависимостей на основе метода наименьших модулей: дис. ... канд. физ.-мат. наук. Екатеринбург, 2019. 148 с.